Is it obligor or instrument that explains recovery rate

Citation for published version:
Yao, X, Crook, J & Andreeva, G 2017, 'Is it obligor or instrument that explains recovery rate: Evidence from US corporate bond', *Journal of Financial Stability*, vol. 28, pp. 1-15. https://doi.org/10.1016/j.jfs.2016.11.005

Digital Object Identifier (DOI):
10.1016/j.jfs.2016.11.005

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Financial Stability

Is It Obligor or Instrument That Explains Recovery Rate: Evidence from US Corporate Bond

Xiao Yao [a,*], Jonathan Crook [b], Galina Andreeva [c]

[a] The Royal Bank of Scotland, 175 Glasgow Road, Gogarburn, Edinburgh, EH12 1HQ, UK

[b, c] Credit Research Centre, The University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK

Abstract: This study investigates the impacts of unobservable firm heterogeneity on modelling corporate bond recovery rates at the instrument level. Based on the recovery information over a long horizon from 1986 to 2012, we find that an obligor-varying linear factor model presents significant improvements in explaining the variations of recovery rates with a remarkably high intra-class correlation being observed. It emphasizes that the inclusion of an obligor-varying random effect term has effectively explained the unobservable firm level information shared by instruments of the same issuer and thus results in an improvement of predictive accuracy of recovery rates. The empirical results show that the latent economic cyclical effects have been well represented by firm level heterogeneity, and strong evidence is presented for the normal distributional assumption of the recovery rates. Finally we demonstrate the choice of recovery rate models may influence portfolio risk with the obligor-varying factor model generating a more right clustered loss distribution than other regression methods on the aggregated portfolio.

Keywords: Unobservable Heterogeneity, Loss Given Default, Portfolio Loss Distribution

JEL classification: **G21 G28**

1. Introduction

Banks that adopt the advanced internal rating based approach (AIRB) are allowed to use their own estimates of risk parameters including probability of default (PD) and loss given default (LGD) to compute the required minimum regulatory capital under Basel II and III (Basel Committee, 2006, 2010). It is necessary for banks and regulators to compute the appropriate amount of capital to be held to protect depositors in the event of insolvency which highlights the importance of an accurate prediction of LGD for the large exposures. Basel II and III consider a single latent factor model proposed by Vasicek (1987, 2002) to be the reference PD model which is based on Merton's model (Merton, 1974) and defined as

$$A = \sqrt{w}X + \sqrt{1 - w}x$$

where the asset value A is assumed to be dependent on a systematic default risk factor X and

* Corresponding author, E-mail : yaoxiao18@gmail.com

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Royal Bank of Scotland. The authors would like to thank the helpful comments of the anonymous referees and of the editor.

an idiosyncratic risk factor x. Vasicek's model assumes that the default contagion effect exists across debt instruments that are dependent on a common single latent factor, which is equivalent to a Probit model with the inclusion of a random effect term. The single factor framework was first adapted to LGD modelling by Frye (2000a) where the collateral value was modelled to be a function of a time-varying systematic risk factor. In his later work (Frye, 2000b) the same methodology was applied to modelling recovery rates of US corporate bonds. The empirical results suggested similar implications that in an economic downturn the expected LGD tended to increase together with an increase in default probability. Dullman and Trapp (2004) developed a joint modelling approach to estimate the correlation between PD and LGD under the single factor framework and to simulate portfolio loss distributions. Rosch and Scheule (2005) proposed a multi-factor framework to model the aggregated annual default rates and recovery rates of corporate bonds jointly with macroeconomic variables incorporated. A similar approach was adopted by Hamerle et al (2006) which incorporated observable explanatory variables such as macroeconomic conditions and individual specific characteristics to estimate the instrument level LGD.

We build our study on previous research to improve model fit under the single factor framework. We do not attempt to identify new determinants of recovery rates. There is a substantial body of work that a reader interested in credit risk determinants can consult. A small selection, which is by no means exhaustive, is provided below as an illustration. Bonfim (2009), using a large sample of Portuguese firms found that both the macroeconomic situation (systemic risk) and firms' financial conditions were important when estimating default probabilities over time. Mora (2015) confirmed the importance of macroeconomic effects on LGD of defaulted US debt securities, but concluded that these differ depending on the industry. For Germany it was found that banks' loss rate was driven by exposure to the global economic situation in addition to the national loss rate, industry, maturity structure and regional factors (Memmel et al., 2015). Frontczak and Rostek (2015) showed that for the retail sector it was important to include collateral and suggested improvements for modelling LGD for this industry. For loans to individuals in Germany, Gürtler and Hibbeln (2013) demonstrated that LGD determinants varied depending on the default/non-default status of loans. Our primary purpose is to investigate the influences of unobservable heterogeneities on modelling corporate bonds recovery rates with the application of a latent factor model.

The major contribution of this study is to incorporate unobservable heterogeneities in empirical recovery rates models by extending the random effect, i.e., the systematic risk factor to multiple levels, and to show its impact on model predictive accuracy. Unlike previous research related to modelling recovery rates using a latent factor model such as Dullman and Trapp (2004), Rosch and Scheule (2005) and Hamerle et al (2006) where the latent factor was assumed to be a time-varying variable representing the general economic condition, we find that by accounting for the firm specific unobservable heterogeneity the single factor model presents improvements in model fit. The empirical evidence presented in our study strongly supports the necessity of including the obligor-varying latent factors. This finding provides a new insight which is that there is a significant amount of valuable account level information about obligors that remains unobserved and can be explained by a random effect term.

A second contribution is that we show that the firm specific intra-class correlation in bond LGD is more significant than other specifications such as seniority and economic levels. The high intra-class correlation suggests that the common accounting information shared by the instruments of the same issuer is largely explained by an obligor level random effect while the benchmarking regression models are not able to account for that. We also examine the predicted latent obligor-varying factors by taking the average of them with respect to each year and find that the aggregated annual obligor-varying latent factors demonstrate similar patterns as the predicted time-varying latent factors. We suggest that the obligor-varying factor specification gives consistently remarkable performances at both instrument and yearly aggregated levels for recovery rates modelling.

A third contribution is to discuss the model impacts under different distributional assumptions for LGD. Few studies have compared the implications of different distributional assumptions for corporate bond recovery rates predictive accuracy. We consider a variety of distributional assumptions for LGD models. We have examined three different benchmarking regression models including linear regression, fractional response regression and inflated beta regression approaches. Linear regression and fractional response regression are well positioned to identify the influential determinants of LGD because of their simplicity and transparency. For example, Acharya et al (2007) studied industry-wide effects on recovery rates and found that an indicator of industry distress influenced instrument recovery rate significantly using a linear regression model. Dermine and Carvalho (2007) proposed a dynamic loan-loss provisioning schedule of SME loans from a Portuguese bank and discussed the effects of covariates on cumulative recovery rates based on a fractional response regression model. Khieu et al (2012) demonstrated that loan characteristics were more significant than other features such as industry and economic conditions prior to default. The beta distribution has been considered to be a promising way to fit recovery rates and was first proposed in Moody's proprietary LossCalc model in Gupton and Stein (2002). Although it has been found that the application of a beta transformation was not as beneficial to recovery rates modelling as expected according to Min and Qi (2009), more recent literature such as Bruche and Aguado (2010) and Jacobs and Karagozoglu (2011) recognized that the beta regression models effectively improved predictions of LGD for corporate bonds. Calabrese (2012) also showed that the inflated beta regression model increased LGD predictive accuracies compared with linear regression and fractional response regression models. Therefore we compare the predictive performances of a collection of regression models including linear regression, fractional response regression and inflated beta regression, and we find that linear regression and fractional response regression are more competitive than inflated beta regression in terms of model fit. Furthermore, we explore three non-linear mixed models including log-normal, logit-normal and inflated beta mixed models where a random effect term is inserted into the linear predictors. Huang and Oosterlee (2011) developed a beta mixed model similar to Bruche and Aguado (2010) to estimate recovery rates of corporate bonds showing a significant improvement of model fit. We follow this idea and propose an inflated beta mixed model by including a random effect term into the predictors of the inflated beta regression, and we show that this inflated beta mixed model yields better predictive accuracies than the inflated beta regression although it is outperformed by the other non-linear specifications such as the log-normal and logit normal. Empirical evidence suggests that both log-normal and logit-normal factor models suffer the same problem that the model fit is highly sensitive to the choice of the perturbation value at the boundaries making them less attractive due to the lack of robustness. By comparing the model performances it is found that none of the non-linear specifications is more favourable than the linear factor model with a normal distributional assumption of the recovery rates.

The last contribution is to analyse the impact of firm level unobservable heterogeneities on portfolio loss distributions at both aggregated and segmented levels. The regulatory focus on accurate and conservative estimation of credit losses as a pre-requisite of financial stability became even more pronounced following the credit crisis. The advanced set of international regulations (Basel Committee, 2010) reconfirmed the importance of regulatory capital, and a new international regulatory body – the Financial Stability Board (FSB) - has been set up to monitor and make recommendations about the global financial system. The following clarification was given in the recent report by FSB's Enhanced Disclosure Task Force: "The regulatory capital framework is designed to ensure that banking organisations maintain capital resources in excess of minimum capital requirements, which reflect both expected and unexpected credit losses. Accounting loan loss allowances are incorporated into the framework and, under IRB approaches, compared to Regulatory EL with the shortfall or excess reflected as an adjustment to capital resources. As a result, regulatory capital EL (under Advanced and Foundation Internal Ratings Based methods) includes prudential floors and downturn estimates…" (EDTF, 2015).

As suggested in the Basel II Accord (Basel Committee, 2005a, 2005b) we implement both AIRB and FIRB approaches to generate portfolio losses where the linear regression and single factor models are employed as AIRB models and the FIRB approach is implemented by specifying a determined value for the bonds with respect to their seniorities provided in Basel II. We find that the aggregated portfolio loss distribution generated by an obligor-varying factor model presents more densely at the right tail than the time-varying factor and linear regression models. This finding is even more important given the recent proposal by the Basel Committee to limit the application of IRB approaches and introduce 'floors to ensure a minimum level of conservatism for portfolios where the IRB approaches remain available' (Basel Committee, 2016). Our results contribute towards the insights into estimation of potential conservative floors. The evidence at the segmented levels is mixed. We notice that the time-varying factor model presents slightly more right clustered loss distributions than the obligor-varying on the senior secured and subordinated bonds but a significant left clustered distribution on the senior subordinated bonds. Another noticeable finding is that the portfolio losses calculated by FIRB approach are significantly lower than the VaR and ES generated by the AIRB approaches. We believe that the LGD specifications of FIRB approach may underestimate the unexpected losses according to the calculated VaR and ES of AIRB approaches for the senior secured and unsecured bonds. However, for the subordinated bonds we find that the FIRB approach provides very close performances to the AIRB models in terms of VaR and ES. The significant discrepancies of portfolio loss distributions between AIRB and FIRB approaches imply that the international regulators should review the LGD requirements of the defaulted corporate bonds to compute the regulatory capital more appropriately, and the financial institutions should be encouraged to develop their internal LGD/RR models to better manage the credit portfolio risk. This is also consistent with the evidence in Jokivuolle and Viren (2013) which showed a joint PD and LGD model produced considerably higher loss risk estimates than a benchmark model with a constant LGD specification although it did not consider the application of heterogeneity.

The remainder of this paper is structured as follows. The next section introduces the models used in the empirical study, and the data is described in Section 3. Section 4 presents the empirical evidence and analysis, and Section 5 discusses the implications of this study for credit risk management. Finally Section 6 concludes the study.

2. Methodology

We first introduce and investigate the specifications under a single factor framework and then we briefly introduce the benchmarking models used in the following empirical study.

2.1 Single factor model

The single factor model in Hamerle et al (2006) is defined as

$$\begin{array}{l} \tilde{y}_i = b_0 + \boldsymbol{\beta}^T \mathbf{x}_i + g_1 Z_t + g_2 e_i \quad t = 1,\dots,T \\ Z_t \sim N(0, 1), \qquad e_i \sim N(0, 1) \end{array}, \tag{1}$$

where Z_t is the random effect term denoting the time-varying systematic recovery risk factor which represents the unobservable heterogeneity of macroeconomic conditions, and e_i is the residual term denoting the idiosyncratic recovery risk factor. Here $\mathbf{x}_i$ is the vector of observable factors and $\tilde{y}_i$ is the transformed recovery rate for instrument i by a logit transformation such that

$$y_i = \exp(\tilde{y}_i) / (1 + \exp(\tilde{y}_i)),$$

where y_i denotes the actual recovery rate. This model can be linked with the specification in Dullmann and Trapp (2004) such that

$$\tilde{y}_i = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i + \sigma\sqrt{\rho} Z_t + \sigma\sqrt{1-\rho}\,\varepsilon_i , \tag{2}$$

where σ is the standard deviation of the recovery rates and ρ is the loading factor denoting the recovery rates correlation between two instruments i and j that default in the same year.

Equation (2) is based on the assumption that the instruments are dependent on a common systematic economic state with respect to the year they default. In this study, we generalize this assumption by specifying that recovery rates of instruments depend on the systematic risk factor with respect to their corresponding obligor or seniority as well as default year. Hamerle et al (2006) find that the inclusion of observable macroeconomic variables renders the systematic risk factor less important statistically. In this study we simply assume recovery rate to follow a normal distribution instead of a logit-normal proposed by Dullmann and Trapp (2004) and Hamerle et al (2006), and we model the actual recovery rate y_i directly such that.

$$\begin{aligned} y_i &= \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i + \gamma_1 Z + \gamma_2 \varepsilon_i \\ Z &\sim N(0,1), \qquad \varepsilon_i \sim N(0,1) \end{aligned} . \tag{3}$$

We show that the normal distributional assumption is more suitable than the other non-normal assumptions in the following empirical analysis. The original time-varying random effect term Z_t in specification (1) will be extended to be obligor and seniority-varying such that Z_k is the *k*-th obligor of the total of K obligors and Z_s is the *s*-th type of the total of S seniorities. The cluster intra-class recovery rate correlation of any two instruments is given as

$$\begin{aligned} Cov(y_i, y_j) &= \sigma^2 \rho = \gamma_1^2 \\ Corr(y_i, y_j) &= \rho = \frac{\gamma_1^2}{\gamma_1^2 + \gamma_2^2} \\ \sigma &= \sqrt{\gamma_1^2 + \gamma_2^2} \end{aligned} . \tag{4}$$

The estimation procedure of this model starts by deriving the conditional probability density function, and then the unconditional density function can be obtained by integrating out the random effects. Conditioning on the realization of Z, y_i follows a normal distribution such as

$$f(y_i \mid Z) = \frac{1}{\sqrt{2\pi}\sigma_Z} \exp\left(-\frac{(y_i - \mu_Z)^2}{2\sigma_Z^2} \right),$$

where

$$\begin{aligned} \mu_Z &= E(y_i \mid Z) = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i + \gamma_1 Z \\ \sigma_Z &= \gamma_2 \end{aligned} .$$

The unconditional probability density function of y_i is then given as the product of the conditional probability density function and the marginal density function of Z such that

$$f(y_i) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_Z} \exp\left(-\frac{(y_i - \mu_Z)^2}{2\sigma_Z^2} \right) d\Phi(Z) ,$$

where $\Phi(Z)$ denotes the cumulative standard normal distribution function. At the final stage the log likelihood function is given by

$$\log L = \sum_i \log f(y_i) .$$

The estimates of the parameters are generated by solving the log likelihood function with standard optimization algorithms. Because of the integral involved in the likelihood function, Gaussian quadrature approximation is adopted before optimizing it.

2.2 Benchmarking regression models

Three linear and generalized linear regression models are selected as the benchmarking regression models including ordinary linear regression, fractional response regression and inflated beta regression. Ordinary linear regression is well known to be stable and transparent for LGD modelling, but it may result in predictions outside of [0, 1] as LGD is defined in this interval. To fix this shortcoming the fractional response regression proposed by Papke and Wooldrige (1996) has been applied to LGD modelling where the dependent variable is defined to be bounded in the open interval (0, 1) by imposing a link function such that

$$E(y \mid \mathbf{x}) = G(\boldsymbol{\beta}^T \mathbf{x}) , \tag{5}$$

where $G(\cdot)$ denotes some link function. We use a logit transformation link function such as

$$G(\boldsymbol{\beta}^T \mathbf{x}) = \exp(\boldsymbol{\beta}^T \mathbf{x}) / (1 + \exp(\boldsymbol{\beta}^T \mathbf{x})) . \tag{6}$$

The first two methods have been extensively investigated in LGD modelling which are commonly used to identify the effects of the determinants of recovery rates. To improve model fit Gupton and Stein (2002) first applied a beta distribution to fitting the irregular LGD distribution. A more promising technique is inflated beta regression proposed by Ospina and Ferrari (2007) to fit the fractional response variables where the dependent variable is defined in the closed interval [0, 1] which can be regarded as a mixture distribution of a beta distribution on (0, 1) and a Bernoulli distribution on bounds 0 and 1. The probability density function is given as

$$bi_{01}(y;p,\gamma,\mu,\phi) = \begin{cases} p(1-\gamma) & if\ y = 0 \\ p\gamma & if\ y = 1 \\ (1-p)f(y;\mu,\phi) & if\ y \in (0,1) \end{cases} . \tag{7}$$

The beta density function $f(y;\mu,\phi)$ is defined as

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} , \tag{8}$$

where μ and ϕ are the mean and precision parameters. To formulate the inflated beta regression model, μ and ϕ can both be reparameterized by link functions such that $\mu = G(\mathbf{x}_\mu)$ and $\phi = H(\mathbf{x}_\phi)$. In this study we only reparameterize μ for comparison purposes. Because of $\mu \in (0,1)$ the link function $G(\mathbf{x}_\mu)$ is chosen to be a logit transformation such as specification (6). This model can also be written in an integrated form as

$$bi_{01}(y;p,\gamma,\mu,\phi) = (p(1-\gamma))^{d}(p\gamma)^{c}((1-p)f(y;\mu,\phi))^{(1-d)(1-c)} , \tag{9}$$

where $d = 1$ if $y = 0$ and $d = 0$ if $y \in (0, 1]$, $c = 1$ if $y = 1$ and $c = 0$ if $y \in [0, 1)$. The expectation of the dependent variable is derived immediately such that

$$E(y) = py + (1 - p)m.$$

The estimates are then derived by maximizing the likelihood function based on (9). More details on fractional response and inflated beta regression models can be obtained from Papke and Wooldrige (1996) and Ospina and Ferrari (2010).

3. Data description

The empirical analysis is built on the recovery rates information from Moody's Ultimate Recovery Database (MURD). This database covers the recovery information of more than 3000 instruments including bank loans, revolvers and corporate bonds. Since we are only interested in the recovery rates of the US corporate bonds, the final sample contains 1413 observations of defaulted instruments observed from 1986 to 2012.

There are five seniorities in our sample: junior subordinated, subordinated, senior subordinated, senior secured and senior unsecured. Each obligor or issuer may issue more than one instrument with varied seniorities. For each instrument the MURD indicates the results of using each of three different methods to calculate recovery rate. Consistent with the literature (Qi and Zhao, 2011), for each instrument we use the method that is recommended by Moody's analysts. This is not the same method for each instrument. In MURD only instrument level information is given including the debt and recovery process characteristics. We integrate the financial ratios as borrowers' characteristics from Compustat into the data of MURD according to the common firm identifier. Macroeconomic variables are also included and extracted from open sources online to capture features of the economic cycle. A full list of the variables used in this study is given in Appendix A. Both accounting and macroeconomic covariates are incorporated into the sample data one year prior to default.

The instrument characteristics include three variables *Collateral Rank*, *Percent Above* and *Issue Size*. *Collateral Rank* denotes the relative rank of the instrument in relation to the other ones issued by the same obligor. A greater value of *Collateral Rank* means the instrument has a relatively lower rank, indicating that it will be recovered after the instruments with a smaller value of *Collateral Rank* have been recovered. Therefore *Collateral Rank* is expected to have negative effects on recovery rate. *Percent Above* is the percentage of the debt that is senior to the current instrument, implying that a higher *Percent Above* is expected to lead to a lower recovery rate. The last variable is *Issue Size* which denotes the original face value of the instrument. Previous findings in the literature on the effect of debt size are mixed. Dermine and Carvalho (2006) showed a significant negative sign of loan size on the recovery rate, and they suggested that the foreclosure of larger loans tended to be delayed by the bank and thus affected the recovery rates negatively. But Acharya et al (2007) found that the debt size was positively related to the recovery rate. They showed that a larger *Issue Size* indicated greater bargaining power for the obligor in the recovery process, and therefore a higher recovery rate was expected.

The firm level information is characterized by seven accounting ratio variables. *Total Asset* can be interpreted as a proxy for firm size, and its expected effect on recovery rates is unclear so far. We might expect that large companies will have lower probabilities of default and higher recovery rates because they have more resources to liquidate their assets to repay the debts during the bankruptcy process. However Acharya et al (2007) pointed out that a large firm might have more difficulties in the process of debt reorganization with higher bankruptcy costs incurred which tended to show a lower recovery rate. *EBITDA* is a measure of the profitability of a firm. The profitability of an obligor's asset is expected to positively influence the recovery rate. A higher *EBITDA* of an obligor indicates that the firm has more cash to cover its debt and is likely to result in a higher recovery rate. *Leverage* is widely studied in the literature with controversial findings. Dwyer and Korablev (2009) demonstrated that a higher leverage increased the PD, hence resulted in a lower recovery rate. Also Acharya et al (2007) argued it was rather difficult to anticipate its effect on recovery rates ex ante. They noted that higher *Leverage* could be related to a more dispersed ownership structure that made the recovery process more complicated. On the other hand, Khieu et al (2012) suggested that a higher value of *Leverage* might imply that it was easier to restructure the debt after bankruptcy suggesting a higher recovery rate. Therefore we do not make any ex ante assumption on this variable. *Book Value per Share*, *Debt Ratio* and *Quick Ratio* are all proxies for the potential solvency capability of a firm. With a higher value of *Book Value per Share*, the firm should have more assets to repay their debts and yield a higher recovery rate. *Debt Ratio* is the ratio of current liabilities to long term debt. It is reasonable to assume that for the short term obligations the funds would be withdrawn immediately and debt extension would be problematic, so the obligor with a high *Debt Ratio* would find it more difficult to repay debts. In summary a higher *Debt Ratio* is expected to affect the recovery rates negatively. A higher *Quick Ratio* indicates an obligor has more cash and short term funds to cover its short term liabilities, which influences the recovery rates positively. We also include the *Asset Tangibility* as in Acharya et al (2007), which is defined as the ratio of tangible assets to intangible assets. Since most intangible assets are difficult to liquidate after default, a higher value of *Asset Tangibility* should be associated with a higher recovery rate.

We include four macroeconomic variables to characterize the US economic conditions including *Growth Rate*, *Unemployment Rate*, *T-Bill Rate* and *Default Rate*. *Growth Rate* is defined as the US annual GDP growth rate. *Growth Rate* and *Unemployment Rate* are used to be the proxy indicators for US economic activities. When the economic activity increases recovery rate should go up accordingly. In other words, recovery rate is expected to be positively correlated with annual GDP *Growth Rate* and negatively with the *Unemployment Rate*. *T-Bill Rate* is the US three months Treasury bill rate as a proxy for the risk free interest rate. A higher *T-Bill Rate* implies a higher cost during the debt recovery process, which subsequently affects the recovery rate negatively (Qi and Zhao, 2011). *Default Rate* has been found to be a powerful predictor of recovery rate and has shown a strong positive influence in previous studies (Altman et al, 2006). *Default Rate* is the annual issuer-weighted corporate default rates (Ou, 2013), which is formulated by averaging the multi-year default rates for all ratings as a useful proxy for the expected default risk.

Figure 1 exhibits the distribution of recovery rates in our sample. It is obvious that the observations are clustered at recovery rates equal to 0 and 1. Panels A and B in Table 1 present the summarized statistics of recovery rates by year and seniority respectively. Here we take the average of recovery rates with equal weights for each issue. The annual mean recovery rates present strong cyclicality, which we will further investigate by including the macroeconomic variables in the regression models. Panel B presents the recovery rates breakdown by seniority. On average bonds with higher seniorities have higher recovery rates than the lower seniorities bonds, although the subordinated and senior subordinated bonds have very similar average recovery rates. Figure 2 shows a plot of annual default rates and recovery rates. The aggregated recovery rates have a strong negative correlation with *Default Rate*. We also observe that *Growth Rate* presents a strong positive relationship with annual recovery rates. In contrast, the *T-Bill Rate* tends to move against recovery rates. However, the relationship between *Unemployment Rate* and recovery rates is ambiguous according to Figure 2, although we expect that a negative relationship between *Unemployment Rate* and the annual recovery rates.

<< FIGURE 1 HERE >>

<< FIGURE 2 HERE >>

<< TABLE 1 HERE >>

Table 2 Panel A exhibits the descriptive statistics of the covariates defined above, and Panel B shows descriptive statistics of instrument covariates by seniorities suggesting the existence of seniority heterogeneity. All the variables are measured at the ratio level except *Collateral Rank*. Based on its definition *Collateral Rank* is an ordinal variable, but it is regarded as a numeric variable in this study. As discussed above its value is negatively correlated economically with recovery rate. The two variables *Issue Size* and *Total Asset* are both subjected to a log transformation for scaling. Notice that three variables: *EBITDA*, *Debt Ratio* and *Book Value per Share* have abnormally large standard deviations Note that the mean and median values of these three variables differ from each other significantly. This strongly suggests that there are outlier values of these variables. Here the values that are higher than 90^{th} percentile or lower than 10^{th} percentile are considered to be the outliers. To preserve the variables' distributions outliers are winsorised at the 10th and 90th quantile of the relevant distributions.

<< TABLE 2 HERE >>

4. Empirical results and analysis

We estimate six models including three specifications of a single factor model and three benchmarking models as described in Section 2. We report the parameter estimates as well as the goodness-of-fit for all models on the whole sample. For the single factor model the random effect is specified at three different levels: obligor, seniority and time. Table 3 exhibits the estimated parameters of the covariates and the metrics of model fit of each model, and in addition, the estimates of the intra-correlation for the factor model are also included.

<< TABLE 3 HERE >>

4.1. Covariates interpretation

Instrument characteristics: First, it is noticed that almost all models show significant negative signs on the parameters of the variable *Collateral Rank*. This is expected and is consistent with our previous analysis. The estimate of the parameter on *Collateral Rank* suggests how much recovery rate is expected to decrease on average if the collateral of an instrument is downgraded by one grade. *Percent Above* interprets the relative seniority in the recovery process of an instrument in a similar way with *Collateral Rank*, where a greater *Percent Above* indicates more debt should be recovered prior to recovering the current instrument. All the models except inflated beta regression demonstrate a significant negative sign of *Percent Above*, which is also consistent with our ex ante hypothesis. Note the magnitude of *Percent Above* is around -0.23 for the linear regression, indicating a very strong negative influence on recovery rate. In other words, an upward change of 0.1 for *Percent Above* is expected to lead to a decrease of more than two percentage points' change for the recovery rate on average. For the last instrument characteristic *Issue Size*, we notice that all the linear models exhibit a negative sign except for the inflated beta regression. A negative linear relationship between Issue Size and bond recovery rate conflicts with the finding in Acharya et al (2007) but such a relationship becomes insignificant in other non-linear regression models. Empirical evidence in Khieu et al (2012) found that *Issue Size* had a negative effect on recovery rate but the estimated parameter was not statistically significant. We suggest that the difficulty for banks to foreclose the large size debts places has more influence than the bargaining power of the issuer and subsequently results in a negative effect on recovery rate.

Firm characteristics: All the models give a significant positive sign for the parameter of *Total Asset* which conforms to the argument that debt restructuring following default tends to be processed more quickly by large companies leading to a higher recovery rate than by small companies. According to Khieu et al (2012) creditors were inclined to trust and to accept a restructuring plan from stockholders with more transparent information indicating more advantages for large firms. In terms of *EBITDA* almost all models present a significant positive sign which coincides with our expectations that a firm with better earning ability should be able to yield a higher recovery rate for its instruments. It is noticed that linear regression gives a significant positive sign for the parameter of *Leverage* and all the other models also confirm a positive effect on the recovery rate but they are not significant. Such evidence would benefit from further investigation to explain the influences of firm debt structure. Next we find that *Debt ratio* poses a significant negative influence on the recovery rate according to Table 3. A higher *Debt ratio* indicates the short term creditors dominate amongst in the obligees, and the short term obligees would prefer to withdraw their funds immediately rather than an extension. Therefore, it is reasonable to observe a negative relationship between *Debt ratio* and recovery rate which is consistent with our finding. As expected the estimated parameter on *Asset Tangibility* is highly significant statistically according to the obligor-varying factor model but other model exhibit mixed evidence. Both *Book Value per Share* and *Quick Ratio* appear to be insignificant for all models, indicating that they are influential determinant of recovery rates.

Macroeconomic variables: We find *Growth Rate* is positively correlated with recovery rate while *Default Rate* has a strongly negative influence as observed in Figure 2. But only the obligor-varying factor model gives a significant parameter estimate on *Growth Rate. T-Bill Rate* also shows a significant negative sign as expected. It conforms to economic intuition that a lower interest rate reduces the cost for an obligor to refinance and restructure its debt leading to a higher recovery rate. The only unexpected result is the significant positive sign of the parameter on *Unemployment Rate*. Economic intuition suggests that a higher *Unemployment Rate* means a more depressed economy where the recovery rate tends to be lower. Considering this variable has s rarely been used in previous research, we believe it might be better explained in further study.

4.2. Goodness of fit

In Table 3 the goodness of fit of all models are presented as well. It shows that the obligor-varying single factor model fits the sample much better than the other methods. It is also noticed that the obligor-varying factor model yields an outstanding model fit with the R^2 of 0.8967. Meanwhile the time-varying factor model also presents better model fit than the other benchmarking regression models. However, the seniority-varying factor model does not demonstrate any improvement in terms of the measure of fit.

We suggest that the improvements of model fit for the single factor model are caused by the inclusion of a random effect which effectively explains unobservable heterogeneity. Notice that the obligor specific intra-class correlation is 0.8308, which is significantly higher than the seniority and time specific levels. This emphasizes that the instruments issued by the same company share a large amount of unobservable common characteristics represented by the high intra-class correlation, and the inclusion of an obligor-varying random effect explains such variations with a significant improvement of model fit. In contrast, the seniority specific intra-class correlation is rather small indicating that the seniority specific unobservable heterogeneity has been sufficiently explained by the instrument characteristics included in this study as shown in Table 2 Panel B.

Among the regression models, fractional response regression shows slightly better model fit compared with the ordinary linear regression, which is consistent with Qi and Zhao (2011) and Khieu et al (2012). Table 3 shows that inflated beta regression model has a lower fit compared with the linear regression model, which conflicts with the findings in Calabrese (2010). One possible explanation is that although the inflated beta regression model accommodates modelling the clustered samples on the boundaries 0 and 1, it is not able to separate the zero or one LGD cases from the remaining cases accurately. In fact it is not unexpected to observe the relatively disappointing performances according to Qi and Zhao (2011), which suggested that the bi-modal distribution should be of secondary concern in LGD modelling. They found that using a beta transformation did not necessarily render a better model fit for a linear regression model. Our finding is also consistent with the latest version LossCalc v3.0 (Dwyer and Korablev, 2009) where the proposed beta transformation in LossCalc v2.0 (Gupton and Stein, 2002) was dropped out and an identity link function was adopted with more model transparency. Dwyer and Korablev (2009) also justified the linear link function specification and showed a clearly linear relationship by grouping the instruments into ten buckets and examining the average actual and predicted LGD of each bucket.

4.3. Unobservable heterogeneities

Our study shows strong evidence for the presence of instrument and macroeconomic characteristics affects in all specifications. Another finding is that with the presence of a time-varying random effect, the macroeconomic variables become less significant indicating that the inclusion of time-varying latent factors weakens the importance of the observable macroeconomic covariates. However, it exhibits mixed results for the change of significance of accounting ratios to be seen by comparing the estimates of obligor-varying factor model and linear regression model in Table 3. In the following we further investigate the influences of unobservable heterogeneities by examining the recovery rates intra-class correlation of single factor models.

The estimates of intra-class correlation r and volatility of recovery rate s in the single factor models are exhibited in Table 3 as defined in Equation (4). First, it is clear that the firm specific intra-class correlation is significantly higher than that of the seniority and time specific levels. We suggest the instruments of the same issuer are highly correlated and the unobservable firm level information is effectively explained by the obligor-varying random effect. It is also straightforward to interpret the low correlation at the seniority level because the instruments with the same seniority do not necessarily demonstrate many common debt characteristics. Furthermore, instruments that defaulted in the same year can be considered to experience the same economic conditions, where the recovery rate correlation is higher than that at the seniority level but still much lower than that at the obligor level. Notice that the estimated volatility of the obligor-varying factor model is 0.3521, which underestimates the historical volatility of 0.3915 as shown in Table 1 but it is the closest estimate compared with the other random effect models.

To examine if the observable covariates can be completely replaced by the random effect defined at the same level, we perform an additional test on three restricted single factor models: the obligor-varying random effect is included with the firm accounting ratios excluded; the seniority-varying random effect is included with the instrument characteristics excluded; and the time-varying random effect is included with the macroeconomic variables excluded. The single factor models with all covariates included are referred to unrestricted models. The estimates of the restricted models are reported in Table 4. We notice that both the magnitudes and signs of the estimates of restricted models are not significantly changed compared with the unrestricted models, indicating that the estimates of factor models are robust enough. It is shown that the obligor-varying random effect can almost replace the effects of observable firm characteristics, because the inclusion of observable firm accounting ratios does not provide any significant improvements in the measure of fit including AIC, BIC and R^2. Our test results suggest that it is sufficient to explain the unobservable firm characteristics by accounting for the firm specific heterogeneity which makes observable accounting ratios almost replaceable. We also find similar evidence for the restricted time-varying factor model, which exhibits very close model fit to the related unrestricted model. The restricted time-varying factor model also shows that the economic cyclical effects can be sufficiently explained by the latent time-varying factors. In fact, our finding is consistent with Rosch and Scheule (2005) which showed that the inclusion of macroeconomic variables decreased the significance of the systematic risk factors. In contrast, the inclusion of a seniority-varying random effect can not replace the observable debt characteristics where the model fit measurements deteriorate significantly when the instrument characteristics are excluded. In fact Table 3 Panel B shows that the instrument characteristics are correlated with bond seniorities suggesting that the seniority heterogeneities have been sufficiently represented by those characteristics.

<< TABLE 4 HERE >>

Finally, to check statistical evidence on the inclusions of latent factors we use the Bayes factor by following the method adopted in Duffie et al (2009), which is represented as twice the differences of the log likelihood between the model with random effect (single factor model) and the null model (ordinary linear regression model). According to the literature cited in Duffie et al (2009), a value of Bayes factor greater than 2 indicates positive strong evidence to include the random effect, and a value greater than 10 indicates very strong evidence. We find that the Bayes factor is 1267.3 for the obligor-varying single factor model and 71.1 for the time-varying model, which provides very strong evidence in favour of including the obligor and time specific latent factors. However, the seniority-varying factor model shows a Bayes factor of 0.3, which indicates it is unnecessary to include the random effect at the seniority level. Such evidence further explains why the model fit can be improved when the obligor and time-varying random effects are included.

4.4. Non-linear factor models

We have discussed the linear factor models based on the normal distributional assumption of recovery rates. To investigate the robustness of this assumption we compare goodness of fit of models based on alternative distributional assumptions about the single factor recovery rates models presented above. We consider three distributional assumptions here: log-normal, logit-normal and inflated beta distribution. The log-normal and logit-normal distributional recovery rates factor model were proposed by Pykhtin (2002) and Dullmann and Trapp (2004) respectively, and the specifications are given as follows.

Log-normal:

$$\ln(y_i) = b_0 + \boldsymbol{\beta}^T \mathbf{x}_i + g_1 Z + g_2 e_i . \tag{10}$$

Logit-normal:

$$\ln\left(\frac{y_i}{1 - y_i}\right) = b_0 + \boldsymbol{\beta}^T \mathbf{x}_i + g_1 Z + g_2 e_i . \tag{11}$$

The estimation procedure is similar to the linear single factor model, where the conditional marginal distribution of recovery rate can be obtained by using the change-of-variable technique. Additionally we formulate the generalized inflated beta model by inserting the random effect term into the linear predictors of the mean parameter as

$$m_Z = G(\mathbf{x}_m + l Z), \tag{12}$$

where Z is the random effect term defined as a standard normal variable with the corresponding scale parameter l. The specification (12) follows the work of Huang and Oosterlee (2011), where beta regression was generalized by including the random effect term into the predictor of expectation m_Z. They suggested that the inclusion of a random effect improved the beta regression model significantly according to the log-likelihood ratio test. However, they did not include any observable covariates in their empirical study and the boundaries 0 and 1 were not defined. Therefore we combine the ideas from Huang and Oosterlee (2011) and Calabrese (2012) by including the random effect term into the linear predictor of inflated beta regression formulating an inflated beta mixed model. This model can be estimated by adopting a similar method to Huang and Oosterlee (2011) where the marginal likelihood can be derived by using the conditional probability density function and integrating out the random effect.

Note that under the log-normal specification RR=0 is undefined and for the logit-normal specification, both RR=0 and RR=1 are undefined. Therefore, a small positive perturbation value t is applied to transform the 0 and 1 to t and $1 - t$ in the implementation of these two models. However, it is rather tricky to select the optimal t for the transformation. Qi and Zhao (2011) have conducted a detailed experiment to investigate sensitivities of t to both in-sample and out-of-sample performances from 1e-11 to 0.5, and they find that the inverse Gaussian regression presents the best in-sample and out-of-sample predictive accuracies when $t = 0.05$.

We argue that the selection of t should not influence the distribution of recovery rates. The 10-th percentile of recovery rates in our data set is 0.0055, and thus we choose 0.001 as the optimal value because when t becomes smaller, the fitted recovery rates deviate from the actual values dramatically. The model fits of three non-linear factor models: log-normal, logit-normal and inflated beta are reported in Table 5. The models have been estimated for the whole sample. We find that all of the models demonstrate the best performances when an obligor-varying latent factor is specified, and the logit-normal factor model shows close model fit compared with linear factor models shown in Tables 3 and 4. Similar to the linear factor model, the model fits of all non-linear factor models deteriorate significantly when the random effect is specified at seniority or time level. In fact we find that both log-normal and logit-normal factor models are highly sensitive to the choice of t implying their unreliable performances. In terms of the generalized inflated beta models, the obligor-varying factor model gives the best model fit while the seniority-varying and time-varying models do not present any improvements. In general the empirical evidence suggests that none of the non-linear factor models outperforms the linear factor model. This is strongly in favour of a linear relationship under the normal distributional assumption for the recovery rate model.

<< TABLE 5 HERE >>

4.5. Out-of-sample prediction

We further examine the out-of-sample predictive performance the linear factor models and compare that with three benchmarking regression models including linear regression, fractional response regression and inflated beta regression in terms of R^2, MAE and RMSE. First we randomly generate a hold-out sample from the whole sample data. However, given that the latent factors are unobservable in the testing set for the factor model, it is necessary to design the experiment more carefully. We notice that the individuals of the same group sharing the same latent factors. For example, if there are two instruments i and j issued by the same obligor k, we select the samples such that instrument i enters into the training set while instrument j is included in the testing set. So suppose we have fitted a linear mixed effects model on the training set, then the systematic risk factor u_k that has been estimated corresponding with instrument i can be applied to predict the recovery rate of instrument j in the hold out sample. In other words, we make sure that any instrument in the testing set should have an instrument issued by the same obligor selected in the training set. We apply this rule also to seniority and time strata. To summarize, we randomly divide all the samples into training and testing sets by a stratified sampling method. The strata are defined at obligor, seniority and time levels to be consistent with the random effect definitions. At each stratum approximately 70 percent of the observations are selected into the training set and the remaining observations are placed in the testing set. The summary statistics for the training and testing sets for different strata are given in Table 6 Panel A, and the in-sample and out-of-sample predictive performance are reported in Table 6 Panel B.

<< TABLE 6 HERE >>

For the benchmarking regression models, we notice that at obligor level stratum fractional response regression gives the highest R^2 of 0.4900 and the lowest RMSE and MAE of 0.2784 and 0.2370 respectively. It also outperforms the other two models at the time level stratum although RMSE and MAE both increase. The inflated beta regression model presents similar performance at seniority level compared to factional response regression. Another interesting finding is that the out-of-sample predictive performances at the obligor level stratum are significantly better than the other strata. This may suggest that when the data are sampled with respect to obligor stratum, the instruments in the holdout sample may share the same accounting information with the instruments of the same obligor in the training set. Therefore the model fitted on the training set should give more accurate predictions because the fitted model can be regarded to have obtained more prior knowledge of the testing set samples. Such evidence is demonstrated more clearly in the single factor models. Among the single factor models it is clear to see that the obligor-varying model outperforms the others substantially achieving RMSE of 0.1489 and MAE of 0.0992 on the holdout sample. However, it is noticed that with the inclusion of a seniority or time-varying factor shows no advantages in terms of the out-of-sample predictive accuracy compared with other benchmarking regression models. This is also consistent with the evidence of model fit, which further confirms that the inclusion of an obligor-varying random effect gives advantages in modelling instruments recovery rates. It also shows that single factor models are robust with similar performances presented on both the training and testing sets.

5. Implications for credit risk management

We investigate the impacts of recovery rate models on the portfolio risk and consider three linear models including obligor-varying and time-varying single factor models as well as ordinary linear regression that may be used as AIRB models to estimate the recovery rates. We simulate the loss distributions based on the AIRB models and compare the characteristics of the loss distributions generated by AIRB approaches by examining the loss distribution characteristics with the Foundation Internal Rating Based approach (FIRB). The implementation procedure is defined as follows.

1) Fit the models on the whole dataset, and collect the parameter estimates of all the covariates. For the single factor model, sample the single systematic risk factor Z and the residual term e_i from independent standard normal distributions. For the linear regression, sample the residual term e_i from a normal distribution $e_i : N(0, s_{ols}^2)$ where s_{ols} is the OLS estimate of volatility. Use the parameters estimated in step 1) to calculate the simulated recovery rates for instrument i. For the instrument i and the simulated recovery rate $\hat{y}_i$, the related simulated

LGD is given by $\hat{\text{LGD}}_i = 1 - \hat{y}_i$.

2) Set the default indicator d_i as $d_i = 1$ for the instrument that defaulted, and assume the exposure at default (EAD) of all instruments equals 1 for simplicity. Calculate the loss rates at the m-th iteration as $L_m = \frac{1}{N}\sum_{i=1}^{N} d_i \hat{\text{LGD}}_i$ since all the instruments in our sample have defaulted.

3) Repeat the above procedures M times and formulate a simulated loss rates distribution.

Here we consider three characteristics including value-at-risk (VaR), expected shortfall (ES) and expected loss (EL) where the last one is defined as the average of loss rates. The definitions of VaR and ES are given in Appendix B. According to the Basel II Accord (Basel Committee, 2005a, 2005b), under the FIRB approach the LGD of each senior unsecured bond is assigned as 0.45 and a subordinated bond is assigned a value of 0.75. Considering that there are five different seniorities in our sample data, we merge the three types of bonds "Junior subordinated bond", "Subordinated bond" and "Senior subordinated bond" as a general type "Subordinated bond" for simplicity, and we keep the other two types "Senior secured bond" and "Senior unsecured bond". The descriptive statistics of LGD with respect to the new categories are given in Table 7. For the LGD of senior secured bonds, banks need to calculate the exposure value after risk mitigation which is not available in MURD. Therefore, we use the historical average LGD of the senior secured bonds in our sample which is 0.3708 in Table 7. We examine the portfolio loss distributions at both aggregated and segmented levels, where the aggregated portfolio refers to the whole sample and the segmented portfolio is given by segmenting the whole sample with respect to the seniorities defined above.

<< TABLE 7 HERE >>

Figure 3 shows the comparisons of loss distributions at both aggregated and segmented levels. Related estimates of VaR, ES and EL are reported in Panels A to D of Table 8, and the t-test statistics to compare EL and ES at aggregated level are provided in Panel E of Table 8. Note that the LGD is a determined value for each instrument under the FIRB approach. The loss rates calculated by the FIRB approach for each segment are just the same as the regulatory values and the aggregated portfolio loss rate is 0.5163. Meanwhile notice that VaR and ES of loss distributions given by the FIRB are generally lower than that of the other AIRB approaches except for the subordinated bonds. We suggest that the FIRB approach may underestimate the extreme losses under a serious economic downturn. The loss distributions generated by linear regression are more concentrated than that of the linear factor models implying that the linear regression model is unable to capture the tail losses, which is clearly undesirable. For the aggregated portfolio the obligor-varying factor model obtains a more right clustered distribution than that of the time-varying factor model according to Panel A of Figure 3. Table 8 Panel E shows that at both

0.05 and 0.01 levels the obligor-varying factor model yields a significantly higher ES and EL compared with the other AIRB and FIRB approaches, suggesting that there are more extreme losses discovered by the obligor-varying model under a severe economic downturn. Similar evidence is found for the senior unsecured bonds based on Panel C, where obligor-varying model generate a higher frequency of tail losses. However, for both senior secured and subordinated bonds, the time-varying model gives greater values of VaR and ES than the obligor-varying model although the differences are not very significant according to Panel B and D. We also find that the LGD specification of the subordinated bonds under FIRB approach is 0.75, which is quite close to the VaR and ES calculated by the AIRB models. We suggest that the LGD specification of subordinated bonds under the FIRB approach is reasonable to buffer the potential unexpected losses, but for the bonds of higher seniorities including senior secured and unsecured bonds FIRB approach may underestimate the tail risk according to the VaR and ES given by AIRB approaches.

<< TABLE 8 HERE >>

<< FIGURE 3 HERE >>

Our results have several implications for regulators who wish to minimise the impact of financial distress on depositors. First, to understand the drivers of LGD and so the factors that affect the optimal amount of capital that lenders should have to protect depositors in the event of severe stress, regulators need to concentrate on developing a deeper understanding of the characteristics of obligors rather than of time varying factors that lead to stress. This is indicated by the much larger proportion of the variance in unexplained variation in LGD due to obligor's specific random effects than to time varying random effects as seen in Table 3.

Second, to predict recovery rates for corporate bonds most accurately in the computation of regulatory capital and economic capital for an existing portfolio an obligor-varying single factor model in addition to instrument level, firm level and macroeconomic observable variables should be considered by regulators and firms. This gives more accurate predictions than models that omit random effects or that include either seniority or time varying random effects instead of random effects specific to the obligor.

Third, the simulation results of Table 8 show that the VaR(0.01) values of the loss distribution due to uncertainty over the variation in LGD that is not explained by the covariates and that is specific to obligors suggest that the implied FIRB LGD parameter in the Basel Accords is too low. That is, randomness in the obligor random effect leads to a VaR (0.01) above the FIRB value.

Fourth, stress testing typically takes the form of computing the amount of regulatory capital needed in the case of a given stress scenario where the scenario is described in terms of macroeconomic variables like those included in our models. But the importance of the random effects, as shown in our results, suggests that these also need to be simulated to get a more accurate assessment of the VaR.

Fifth, looking at the time varying random effects VaR values, if a period of stress occurs and is due not to severe changes in observables like growth rate, the T-Bill rate, default rate or unemployment rate but instead due to, or described by, random changes in other and unobserved phenomena, the FIRB LGD parameters are again too low. For example in Table 8 panel A the VAR (0.01) for the time varying factor model is 0.5823 under AIRB, but 0.5163 under FIRB. The same relative values hold when the regulatory capital is computed separately for each level of seniority. Given that the inclusion of time varying random effects increases model fit and the proportion of the total error accounted for this random term is significant, the aspects of the economy we do not understand may well be causes of major changes in required LGD.

6. Concluding remarks

Unobservable heterogeneity has been well investigated in PD modelling where default risk is assumed to be correlated across different instruments and firms are dependent on a common risk factor. In recovery rates models a latent time-varying systematic risk factor is commonly incorporated to explain the economic cyclical effects on recovery risk. In this paper we investigate the impact of firm specific heterogeneity on modelling US corporate bond recovery rates and make the following contributions to the literature.

First we place the emphasis on the inclusion of firm heterogeneity in modelling instrument level recovery rates. By specifying the random effect at the obligor level, the single factor model shows a substantial improvement in model fit compared with the time-varying factor model and other traditional regression models. We suggest that the main reason is that the unobservable obligor information is well explained by accounting for the firm specific heterogeneity. Unlike Hamerle et al (2006) who argued that more important observable variables were needed to explain the variations of LGD, our findings suggest that the reason why the variations of recovery rates can not be explained adequately is caused by the absence of heterogeneity instead of the absence of other relevant observable determinants.

Next we show that the specification of the normal distributional assumption is more appropriate than the other non-normal distributional assumptions for the errors. Our finding is consistent with Dwyer and Korablev (2009) which has shown that it is reasonable to assume a linear relationship for the recovery rate and its determinants. For the benchmarking regression models it is noted that fractional response regression gives marginal advantages to linear regression and inflated beta regression in terms of model fit and predictive accuracies. However, the linear factor model is more robust with better model fit than the other non-linear specifications. We compare three other distributional assumptions on the factor models, and find that only the logit-normal specification gives a comparable model fit. Both log-normal and logit-normal factor models are extremely sensitive to the choice of the perturbation value at boundaries 0 and 1. The inflated beta mixed model proves to be more advantageous than the inflated beta regressions, but they are not comparable with the linear models. We believe a linear specification is a suitable choice for bonds recovery rates modelling.

Furthermore we examine the in-sample and out-of-sample prediction of linear factor models and benchmarking regression models, and we find that obligor-varying factor model consistently outperforms other regression models in terms of predictive accuracy. Similar to the finding of model fit the inclusion of a seniority- or time-varying random effect makes little contribution to out-of-sample predictive accuracy compared to a linear regression. Empirical evidence also finds that single factor models present robust performances on both training and holdout samples.

Finally we investigate the impact of our models on credit risk management by comparing the simulated loss rates distributions generated by the AIRB approaches represented by single factor models and linear regression and the FIRB approach. We find that under the FIRB approach the aggregated portfolio loss is seriously underestimated. For the segmented portfolios the LGD specification of subordinated bonds under FIRB is rather close to the estimates of VaR and ES of the AIRB models and is appropriate for the loss predictions. But for the bonds with higher seniorities the FIRB approach may underestimate the unexpected losses based on our simulation results. We also find that both obligor-varying and time-varying models provide more frequent extreme losses than the linear regression method for both aggregated and segmented portfolios. We suggest the LGD specifications under FIRB approach may underestimate the potential unexpected losses, especially for the bonds with high seniorities.

One caveat in our study is that the default and recovery risk correlation is not taken into consideration in our modelling framework because of the limit of the data. We believe that with the incorporation of a default risk model, we expect to observe even more interesting evidence of the impact of the firm heterogeneity on the credit portfolio losses.

References

Acharya, V. V., Bharath, S.T. & Srinivasan, A., 2007. Does industry-wide distress affect defaulted firms? Evidences from creditor recoveries. Journal of Financial Economics 85, 787-821.

Altman, E. D., 2006. Default recovery rates and LGD in credit risk modeling and practice. Working paper.

Basel Committee on Banking Supervision, 2005a. Guidance on paragraph 468 of the framework document.

Basel Committee on Banking Supervision, 2005b. An explanatory note on the Basel II IRB risk weight functions.

Basel Committee on Banking Supervision, 2006. Basel II: International convergence of capital measurement and capital standards: A revised framework.

Basel Committee on Banking Supervision, 2010. Basel III: Strengthening the resilience of the banking sector.

Basel Committee on Banking Supervision, 2016. Reducing variation in credit risk-weighed assets constraints on the use of internal model approaches.

Bastos, J. A., 2010. Forecasting bank loans loss-given-default. Journal of Banking and Finance 34(10), 2510-2517.

Bluhm, C., Overbeck, L. & Wagner, C., 2003. An introduction to credit risk modeling. Chapman & Hall/CRC.

Bonfim, D., 2009. Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics, Journal of Banking & Finance, 33(2), 281-299.

Bruche, M. & Gonzalez-Aguado. Carlos., 2010. Recovery rates, default probabilities, and the credit cycle. Journal of Banking and finance 34, 754-764.

Calabrese, R., 2010. Predicting bank loan recovery rates in a mixed continuous-discrete model. Working Paper.

Chava, S., Stefanescu, C. & Turnbull, S., 2011. Modeling the loss distribution. Management Science 57(7), 1267-1287.

Dermine, J. & Neto De Carvalho, C., 2006. Bank loan losses-given-default: A case study. Journal of Banking and Finance 30, 1219-1243.

Dermine, J. & Neto De Carvalho, C., 2008. Bank loan-loss provisioning, central bank rules vs. estimation: The case of Portugal. Journal of Financial Stability 4, 1-22.

Dullmann, K. & Trapp, M., 2004. Systematic risk in recovery rates-an empirical analysis of US corporate credit exposures. Deutsche Bundesbank Discussion Paper.

Duffie, D., Eckner, A., Horel, G. & Saita, L., 2009. Frailty correlated default. Journal of Finance 64(6), 2089-2132.

Dwyer, D. & Korablev, I., 2009. Moody's KMV LossCalc V3.0. Moody's Analytics.

EDTF (Enhanced Disclosure Task Force), 2015. Impact of expected credit loss approaches on bank risk disclosures.

Frontczak, R. and Rostek, S., 2015. Modeling loss given default with stochastic collateral, Economic Modelling, 44, 162-170.

Frye, J., 2000a. Collateral damage: A source of systematic credit risk. Risk, 91-94.

Frye, J., 2000b. Depressing recoveries. Risk, 108-111.

Gordy, M., 2003. A risk-factor model foundation for rating-based bank capital rules. Journal of Financial Intermediation 12, 199-232.

Gupton, G. M. & Stein, R. M., 2002. LossCalc™: Model for predicting loss given default (LGD). Moody's KMV.

Gürtler, M. and Hibbeln, M., 2013. Improvements in loss given default forecasts for bank loans, Journal of Banking & Finance 37(7), 2354-2366.

Jacobs, Jr. M. & Karagozoglu, A. K., 2011. Modeling ultimate loss-given-default on corporate debt. Journal of Fixed Income 21(1), 6-20.

Jakivuolle, E., Viren, M., 2013. Cyclical default and recovery in stress testing loan losses. Journal of Financial Stability 9, 139-149.

Khieu, H., D. Mullineaux, D. J. & Yi, H. C., 2012. The determinants of bank loan recovery rates. Journal of Banking and Finance 36, 923-933.

Hamerle, A., Knapp, M. & Wildenauer, N., 2006. Modeling loss given default: A "point in time"-approach. The Basel II Risk Parameters. Springer, Berlin.

Hillebrand, M., 2005. Modeling and estimating dependent loss given default. Risk, 120-125.

Huang, X. & Oosterlee, C. W., 2011. Generalized beta regression models for random loss given default. Journal of Credit Risk 7(4), 1-26.

Memmel, C., Gündüz, Y., and Raupach, P., 2015. The common drivers of default risk, Journal of Financial Stability 16, 232-247.

Merton, R. C., 1974. The pricing of corporate debt: The risk structure of interest rates. Journal of Finance 29, 449-470.

Mora, N., 2015. Creditor recovery: The macroeconomic dependence of industry equilibrium, Journal of Financial Stability 18, 172-186.

Moody's Analytics, 2012. Default & recovery database DRD technical specifications.

Ospina, R & Ferrari, S., 2010. Inflated beta distributions. Statistics Papers 51, 111-126.

Ou, S., 2013. Annual default study: Corporate default and recovery rates, 1920-2012. Moody's Investors Service.

Pykhtin, M., 2003. Unexpected recovery risk. Risk, 74-48.

Papke, L. & Wooldridge, J., 1996. Econometric method for fractional response variables with an application to the 401(K) plan participation rates. Journal of Applied Econometrics 11, 619-632.

Qi, M. & Yang, X., 2009. Loss given default of high loan-to value residential mortgages. Journal of Banking and Finance 33, 788-799.

Qi, M. & Zhao, X., 2011. Comparison of modeling methods for loss given default. Journal of

Banking and Finance 35, 2842-2855.
Rosch, D. & Scheule, H., 2005. A multi-factor approach for systematic default and recovery risk. Journal of Fixed Income 15(2), 63-75.
Tasche, D., 2004. The single risk factor approach to capital charges in case of correlated loss given default rates. Deutsche Bundesbank Working Paper.
Rosch, D. & Scheule, H., 2004. Forecasting retail portfolio credit risk. Journal of Risk Finance 5, 16-32.
Schonbucher, P. J., 2001. Factor models for portfolio credit risk. Journal of Risk Finance 3, 45-56.
Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model. Journal of Business 74(1), 101-124.
Vasicek, O., 1987. Probability of loss on loan portfolio. KMV Corporation.
Vasicek, O. 2002. Loan portfolio value. Risk, 160-162.

Appendix A. List of covariates

Collateral Rank: the rank that the instruments are ranked in relation to each other within the same obligor based on the structure prior to default, where the collateral and instrument seniority are also taken into consideration.

Percent Above: the percentage of debt which is contractually senior to the current instrument.

Issue Size: the face value of the instrument

Total Asset: the total asset value of the obligor.

EBITDA: the earnings before interest, taxes, depreciation and amortization.

Leverage: the ratio of total debts to total assets.

Book Value per Share: the book value of assets scaled by the total outstanding shares.

Debt Ratio: the ratio of current liabilities to long term debt.

Asset Tangibility: the ratio of tangible assets to intangible assets.

Quick Ratio: the sum of cash and short term investment and total receivables divided by the current liabilities.

Growth Rate: the US annual GDP growth rate.

Unemployment Rate: the US annual unemployment rate.

T-Bill Rate: the US three months Treasury bill rate. This is a proxy for the risk free interest rate.

Annual Default Rates: the annual issuer-weighted corporate default rates.

Appendix B. Definitions of Value-at-Risk (VaR) and Expected Shortfall (ES)

Given a confidence level $q \in (0, 1)$, the VaR and ES of a loss distribution are defined as

$$\begin{aligned} \mathrm{VaR}_q(L) &= \min\{l \mid P(L > l) \le 1 - q\} \\ \mathrm{ES}_q &= E(L \mid L > \mathrm{VaR}_q) \end{aligned},$$

where l is the smallest value such that the probability that the loss rate L exceeds l is $1 - q$ at most. Here we use the empirical quantile as the estimate of VaR such that

$$\mathrm{VaR}_q(L) = \hat{L}_q,$$

where $\hat{L}_q$ satisfies that $P(L \ge \hat{L}_q) = 1 - q$. Given the estimate of VaR, $\hat{L}_q$, we can derive the empirical estimate of the ES as follows:

$$\mathrm{ES}_q = \frac{1}{N_q} \sum_{j=1}^{M} L_j I\left[L_j > \hat{L}_q\right],$$

where L_j denotes the loss rate simulated at j-th iteration, and $I[\cdot]$ is an indicator function, and N_q is the count that the loss rate L exceeds $\hat{L}_q$.

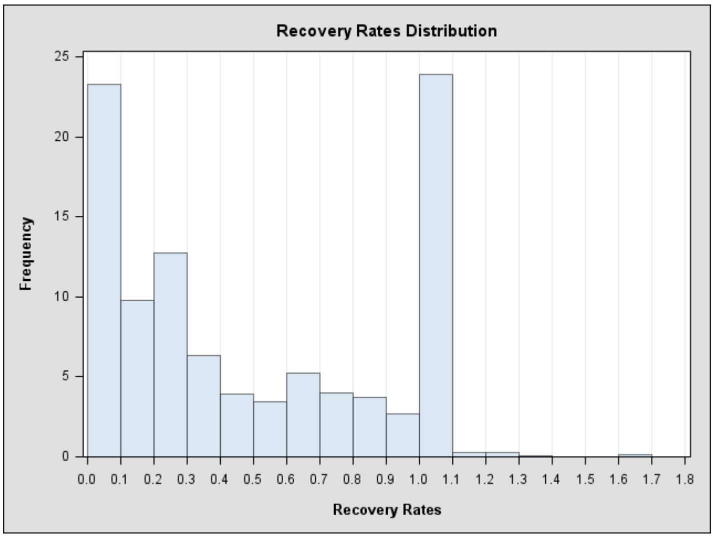


Fig. 1. Distribution of recovery rates

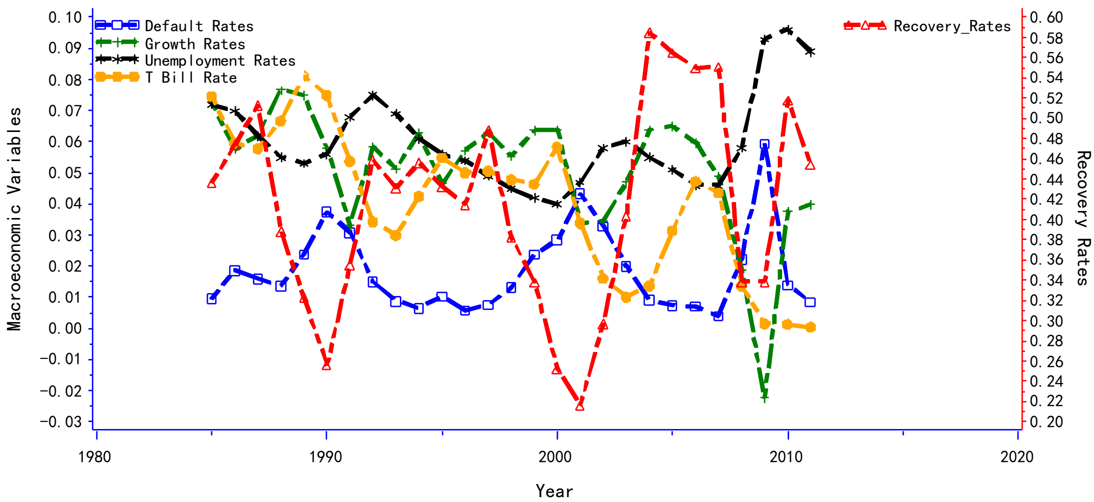
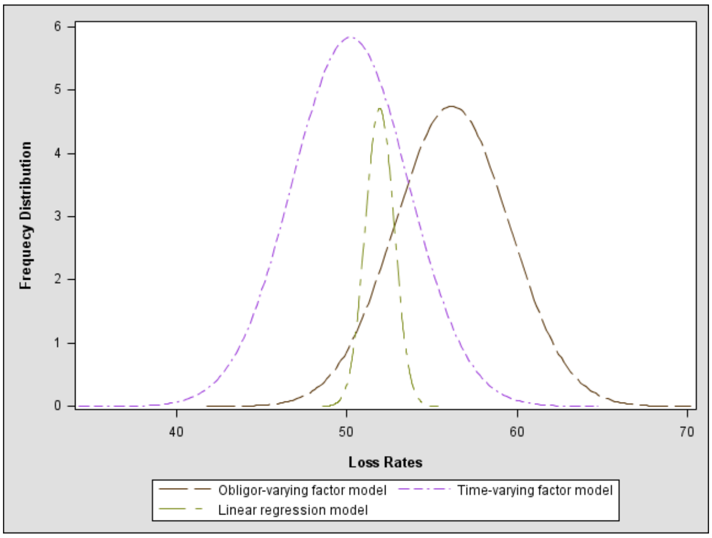
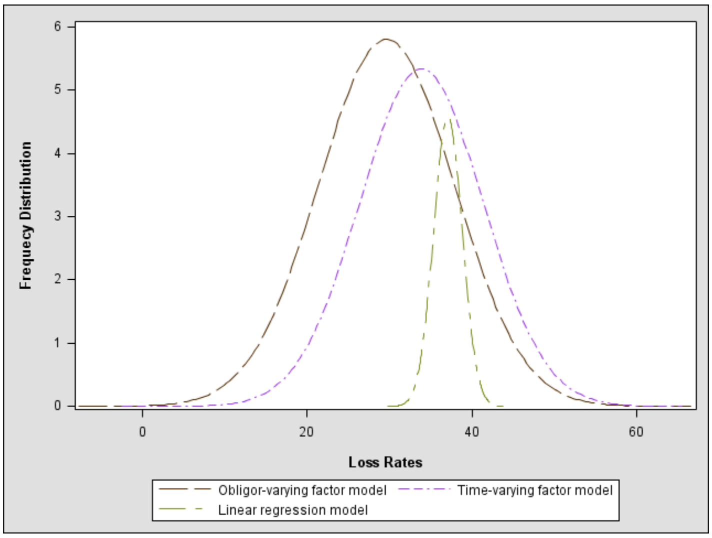


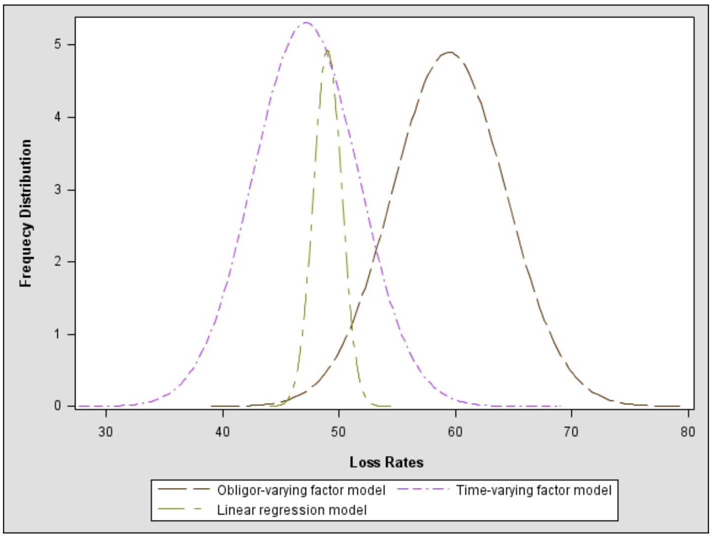
Fig. 2. Plot of macroeconomic variables against recovery rates



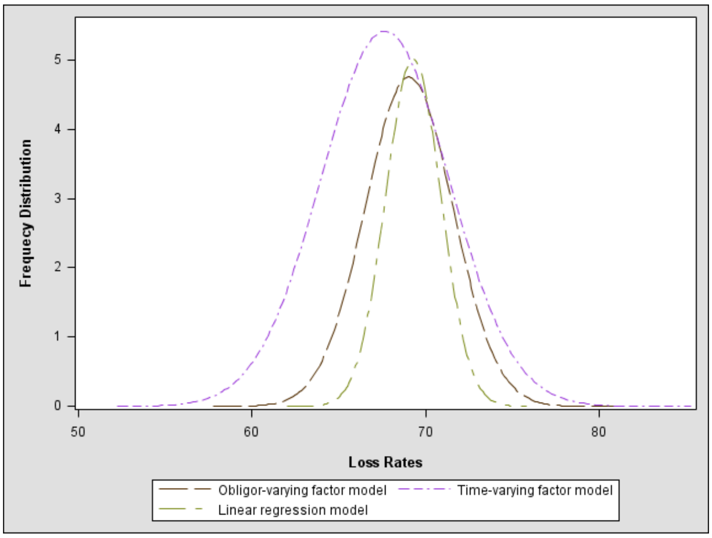
Panel A. Aggregated portfolio



Panel B. Senior secured bonds



Panel C. Senior unsecured bonds



Panel D. Subordinate bonds

Fig. 3. Plot of simulated loss distributions

Table 1. Descriptive statistics of recovery rates

Table 1 presents the descriptive statistics of recovery rates by year and seniority respectively. For each instrument the MURD indicates the results of using each of three different methods to calculate recovery rate and for each instrument we use the method that is recommended by Moody's analysts. There are three methods provided in MURD to calculate RR for each instrument. Discount_Settlement_Total: The nominal settlement recovery amount discounted back from each settlement instrument's trading date to the last date cash paid of the individual defaulted instruments, using the defaulted instrument's effective interest rate; Discount_Liquidity_Total: The nominal liquidity recovery total discounted back from each settlement instrument's trading date to the last date cash paid of the individual defaulted instruments, using the defaulted instrument's effective interest rate; Discount_Trading_Price: The trading price nominal recovery value discounted from the trading date to the instrument's last date cash paid using the effective interest rate of the pre-defaulted instrument.

Panel A. Recovery rates breakdown by year

	No.	Mean	Std	Min	Max
1986	3	0.1436	0.1030	0.0272	0.223
1987	12	0.6195	0.2585	0.3185	1
1988	17	0.7850	0.3998	0.048	1
1989	14	0.1438	0.1695	0	0.4731
1990	65	0.2723	0.2849	0	1
1991	78	0.5629	0.3864	0	1
1992	47	0.6267	0.4287	0	1
1993	19	0.4104	0.4035	0	1
1994	13	0.6510	0.4177	0.0843	1
1995	25	0.6128	0.3806	0	1
1996	5	0.2989	0.4113	0.0024	1
1997	25	0.2687	0.1728	0	0.5577
1998	39	0.2977	0.3362	0	1
1999	54	0.3522	0.3558	0	1
2000	139	0.5449	0.4504	0	1
2001	192	0.3019	0.3360	0	1.012
2002	221	0.3840	0.3091	0	1.3691
2003	88	0.6578	0.3511	0	1.1298
2004	38	0.7244	0.4134	0	1.2766
2005	96	0.8026	0.2790	0	1.6978
2006	16	0.6299	0.4310	0	1.1567
2007	15	0.5704	0.4103	0.0024	1.013
2008	100	0.4910	0.3983	0	1.0029
2009	75	0.4903	0.3851	0	1.0373
2010	12	0.6228	0.4678	0.0039	1
2011	4	0.2870	0.2823	0.0455	0.5857
2012	1	0.2705	.	0.2705	0.2705
Total	1413	0.4806	0.3915	0	1.6978

Panel B. Recovery rates breakdown by seniority

	No.	Mean	Std	Min	Max
Senior secured bonds	332	0.6292	0.3688	0	1.1298
Senior unsecured bonds	681	0.5100	0.3813	0	1.0499
Senior subordinate bonds	198	0.3150	0.3617	0	1.6978
Subordinate bonds	174	0.3217	0.3743	0	1.3691
Junior subordinate bonds	28	0.1628	0.2634	0	1.0000
Total	1413	0.4806	0.3915	0	1.6978

Table 2. Descriptive statistics

Panel A presents the descriptive statistics of all covariates in the empirical study which are classified into three categories: instrument characteristics, firm characteristics and macroeconomic variables. I

Panel B shows descriptive statistics of instrument covariates including Collateral Rank, Percent Above and Issue Size by seniority.

Panel A. Descriptive statistics of all covariates

	Mean	Median	Std	Min	Max
		Instrument characteristics			
Collateral Rank	2.1677	2.0000	0.9277	1.0000	6.0000
Percent Above	0.3230	0.2812	0.2894	0.0000	1.0000
Issue Size	2.299E8	1.427E8	3.081E8	100,000	3.987E9
		Firm characteristics			
Total Asset	10140.30	1570.91	20886.31	16.8320	103914
EBITDA	680.8151	55.6090	1920.55	-2439.94	10489
Leverage	0.9945	0.8843	0.4575	0.2893	4.8787
Debt Ratio	19.0926	0.4639	518.8207	0.0436	19455.2
Book Value per Share	1687.89	1.6962	46,726.03	-875083	255000
Asset Tangibility	0.3310	0.1344	0.5004	0.0000	5.1922
Quick Ratio	0.7196	0.5565	0.6549	0.0124	5.9174
		***Macroeconomic variables* (%)**			
Growth Rate	5.1017	5.8080	1.6635	-2.2237	7.6852
T-Bill Rate	4.2160	4.3600	1.9791	0.0500	8.1100
Default Rate	2.4307	2.3770	1.2968	0.3980	5.9340
Unemployment Rate	5.1183	4.7000	0.9521	4.0000	9.6000

Panel B. Descriptive statistics of instrument covariates by seniority

	Mean	Median	Std	Min	Max
		Collateral Rank			
Senior secured bonds	1.5843	2.0000	0.5834	1.0000	4.0000
Senior unsecured bonds	2.2599	2.0000	0.9904	1.0000	6.0000
Senior subordinate bonds	2.4394	2.0000	0.7566	1.0000	5.0000
Subordinate bonds	2.4023	2.0000	0.8183	1.0000	5.0000
Junior subordinate bonds	3.4643	3.0000	0.7927	2.0000	5.0000
		Percent Above			
Senior secured bonds	0.0911	0.0581	0.1401	0.0000	0.7136
Senior unsecured bonds	0.3131	0.3201	0.2451	0.0000	0.9760
Senior subordinate bonds	0.4884	0.5152	0.2518	0.0000	0.9996
Subordinate bonds	0.5385	0.6440	0.3401	0.0000	1.0000
Junior subordinate bonds	0.8053	0.8148	0.1360	0.5491	0.9955
		***Issue Size* (\$*m*)**			
Senior secured bonds	149.14	85.00	217.56	0.1000	1500
Senior unsecured bonds	313.25	200.00	383.66	0.5000	3987
Senior subordinate bonds	179.07	155.28	126.13	0.8000	700
Subordinate bonds	129.72	100.00	172.31	1.4300	1600
Junior subordinate bonds	145.31	102.23	158.16	1.0000	750

Table 3. Estimates of parameters

Table 3 demonstrates the estimated results of regression models applied to the whole sample data. Both *Issue Size* and *Total Asset* are subjected to a log transformation. Here *, ** and *** represent significance at 10%, 5% and 1% level respectively, and t-values of the estimated parameters are reported in parentheses. Note that t values and their corresponding significance levels are not consistent across models due to the change of degree of freedom. For linear regression the OLS estimate of volatility is reported.

	Single factor models			Linear regression	Fractional response regression	Inflated beta regression
	Obligor	**Seniority**	**Time**			
Intercept	0.4174 ***	0.5486 **	0.2775 ***	0.5879 ***	0.2832	-0.6021
	(3.70)	(2.41)	(4.02)	(3.99)	(0.26)	(-1.00)
Collateral Rank	-0.1844 ***	-0.1277 ***	-0.1265 ***	-0.1283 ***	-0.6749 ***	-0.2400 ***
	(-19.95)	(-9.71)	(-9.59)	(-9.7)	(-6.78)	(-5.92)
Percent Above	-0.0804 ***	-0.2139 ***	-0.2263 ***	-0.2272 ***	-1.0901 ***	-0.9740
	(-2.76)	(-4.88)	(-5.55)	(-5.61)	(-3.80)	(-0.47)
Log(Issue Size)	-0.0176 ***	-0.0149 **	-0.0190 ***	-0.0174 **	-0.0817	0.0114
	(-5.04)	(-2.37)	(-3.27)	(-2.95)	(-1.95)	(0.47)
Log(Total Asset)	0.0627 ***	0.0489 ***	0.0414 ***	0.0495 ***	0.2546 ***	0.0882 ***
	(10.13)	(6.67)	(5.58)	(6.96)	(4.93)	(3.16)
EBITDA	0.00002	0.00007 ***	0.00008 ***	0.00006 ***	0.0003 ***	0.00005
	(0.01)	(4.81)	(5.32)	(4.65)	(2.92)	(0.93)
Leverage	0.0429 **	0.0313	0.0059	0.0319	0.1888	-0.0741
	(1.89)	(1.33)	(0.26)	(1.34)	(1.10)	(-0.79)
Debt Ratio	-0.0014	-0.0088 **	-0.0086 ***	-0.0088 ***	-0.0515 **	-0.0226 ***
	(-0.77)	(-3.92)	(-3.93)	(-3.90)	(-2.94)	(-2.65)
Book Value per Share	-0.0006	-0.0004	-0.0004	-0.0005	-0.0022	-0.0016
	(-1.64)	(-1.01)	(-0.92)	(-1.10)	(-0.74)	(-0.98)
Asset Tangibility	0.0686 ***	-0.0162	-0.0148	-0.0161	-0.0702	0.0048
	(3.89)	(-0.86)	(-0.81)	(-0.86)	(-0.53)	(0.07)
Quick Ratio	-0.0188	-0.0186	-0.0097	-0.0190	-0.1060	-0.0543
	(1.30)	(-1.33)	(-0.71)	(-1.34)	(-1.07)	(-1.00)
Growth Rate	0.6413	1.8884	0.0614	1.8886 **	9.1749	4.9603
	(1.30)	(2.08)	(0.02)	(2.08)	(1.37)	(1.35)
T-Bill Rate	-0.0077 *	-0.0456 ***	-0.0365 *	-0.0457 ***	-0.2337 ***	-0.1857 ***
	(-1.68)	(-7.29)	(-1.98)	(-7.27)	(-5.05)	(-6.93)
Default Rate	-0.0220 ***	-0.0437 ***	-0.0345	-0.0432 ***	-0.2017 ***	-0.1146 ***
	(-2.75)	(-4.67)	(-1.33)	(-4.62)	(-3.04)	(-3.00)
Unemployment Rate	0.0490 ***	0.0698 ***	-0.0006	0.0707 ***	0.3679 ***	0.1754 ***
	(4.43)	(6.69)	(-0.02)	(6.80)	(4.46)	(3.97)
s	0.3521 ***	0.3165 ***	0.3264 ***	0.3164 ***		
	(39.48)	(52.97)	(32.97)	(53.16)		
r	0.8308 ***	0.0037	0.1427 ***			
	(73.70)	(0.67)	(3.12)			
-2loglikelihood	-493.3	769.2	678.0	758.2	1620.7	1683.2
AIC	-459.3	803.2	712.0	790.2	1650.7	1719.2
BIC	-391.6	796.6	712.5	790.6	1729.5	1813.8
R^2	0.8967	0.3383	0.4122	0.3461	0.3679	0.2084

Table 4. Estimates of the restricted single factor models

Table 4 presents the estimated results of three restricted single factor models. Obligor: the obligor-varying model with firm characteristics excluded; Seniority: the seniority-varying model with instrument characteristics excluded; Time: the time-varying model with macroeconomic variables excluded. Here *, ** and *** represent significance at 10%, 5% and 1% level respectively, and t-values of the estimated parameters are reported in parentheses. Please note that s and r are defined as equation (4).

	Restricted single factor models		
	Obligor	**Seniority**	**Time**
Intercept	0.9485 ***	-0.3070 *	0.8798***
	(10.12)	(-2.50)	(6.07)
Collateral Rank	-0.1868 ***		-0.1270 ***
	(-21.64)		(-9.75)
Percent Above	-0.0576 **		-0.2216 ***
	(-2.09)		(-5.55)
Log(Issue Size)	-0.0165 ***		-0.0183 **
	(-5.43)		(-3.17)
Log(Total Asset)		0.0284 **	0.0411 ***
		(3.81)	(5.59)
EBITDA		0.00006 **	0.00008 ***
		(3.88)	(5.44)
Leverage		0.0109	0.0070
		(0.43)	(0.31)
Debt Ratio		-0.0104 **	-0.0088 ***
		(-4.36)	(-4.04)
Book Value per Share		0.0007	-0.0004
		(1.57)	(-1.01)
Asset Tangibility		-0.0438 *	-0.0145
		(-2.21)	(-0.80)
Quick Ratio		0.0005	-0.0087
		(0.04)	(-0.64)
Growth Rate	0.8974	3.7749 **	
	(1.30)	(3.90)	
T-Bill Rate	-0.0169 ***	-0.0366 ***	
	(-3.75)	(-5.69)	
Default Rate	-0.0151 **	-0.0127	
	(-2.10)	(-1.30)	
Unemployment Rate	0.0346 ***	0.0985 ***	
	(3.86)	(8.291)	
s	0.3408 ***	0.3324	0.3407 ***
	(59.93)	(2.08)	(30.42)
r	0.8166 ***	0.1628	0.2147 ***
	(88.20)	(1.03)	(4.56)
-2loglikelihood	-458.4	978.6	685.6
AIC	-438.4	1006.6	711.6
BIC	-398.5	1001.1	728.4
R^2	0.8948	0.2482	0.4138

Table 5. Model fit of non-linear single factor models

Table 5 shows the model fit of three non-linear single factor models. Three different non-normal distributional assumptions are considered to including log-normal, logit-normal and inflated beta distributions. The random effect is specified at obligor, seniority and time levels as above. Note that a small positive perturbation value t is applied to transform the 0 and 1 to t and $1 - t$ in the implementation of log-normal and logit-normal factor models because the boundary points are not defined. We find that the model fit of these two models are highly sensitive to the choice of t, and we choose 0.001 as the optimal value. R^2 is reported as the measure of model fit.

	Random effect	**-2loglikelihood**	**AIC**	**BIC**	**R^2**
Log-normal	obligor	-320	-286	-218.2	0.4955
	seniority	671.4	705.4	698.8	0.0122
	time	626.3	660.2	682.9	0.116
Logit-normal	obligor	-412.3	-408.9	-402.1	0.8151
	seniority	813.7	810.3	808.1	0.0524
	time	791.3	787.9	788.6	0.1132
Inflated beta	obligor	298.8	392.8	580.1	0.5994
	seniority	1211.3	1305.3	1286.9	0.3309
	time	1117.4	1211.4	1272.3	0.3726

Table 6. Out-of-sample prediction

Panel A shows the settings of training and testing sets for out-of-sample prediction. The training and testing sets are divided based on stratified sampling method with strata defined at obligor, seniority and time levels. At the obligor stratum, any instrument in the testing set should have an instrument issued by the same obligor selected in the training set. This rule is also applied to seniority and time strata.

Panel B presents both in-sample and out-of-sample prediction performances including three performance metrics: R^2, RMSE and MAE. For single factor models, the sample strata are consistent with the random effect levels for out-of-sample predictions.

Panel A. In-sample and Out-of-sample set up

	In-sample		Out-of-sample	
Strata	**Obligors**	**Instruments**	**Obligors**	**Instruments**
Obligor	398	1037	144	376
Seniority	352	991	196	422
Time	356	1002	197	411

Panel B. Model prediction performance

	Sampling Strata	**In-sample**			**Out-of-sample**		
		R^2	RMSE	MAE	R^2	RMSE	MAE
Single factor model	Obligor	0.8962	0.1243	0.0842	0.8640	0.1489	0.0992
	Seniority	0.3648	0.3111	0.2546	0.2994	0.3288	0.2710
	Time	0.4030	0.3010	0.2452	0.4156	0.3022	0.2424
Linear regression model	Obligor	0.3162	0.3231	0.2673	0.4703	0.2837	0.2429
	Seniority	0.3543	0.3137	0.2579	0.3112	0.3260	0.2691
	Time	0.3380	0.3170	0.2633	0.3590	0.3165	0.2578
Fractional response regression model	Obligor	0.3348	0.3187	0.2613	0.4900	0.2784	0.2370
	Seniority	0.3776	0.3080	0.2501	0.3380	0.3196	0.2604
	Time	0.3577	0.3122	0.2562	0.3823	0.3107	0.2511
Inflated beta regression model	Obligor	0.3142	0.3196	0.2702	0.4042	0.3118	0.2662
	Seniority	0.3649	0.3111	0.2577	0.3402	0.3190	0.2672
	Time	0.3475	0.3147	0.2639	0.3647	0.3151	0.2595

Table 7. Summarized statistics of LGD for aggregated and segmented portfolios

Table 7 presents the summarized statistics of LGD for aggregated and segmented portfolios. The aggregated portfolio is represented by the whole sample and segmented by seniority. The three types of bonds "Junior subordinated bond", "Subordinated bond" and "Senior subordinated bond" are merged as a general type "Subordinated bond" for simplicity, and the other two types "Senior secured bond" and "Senior unsecured bond" are kept.

	No.	**Mean**	**Std**
Senior secured bonds	332	0.3708	0.3688
Senior unsecured bonds	681	0.4900	0.3813
Subordinated bonds	400	0.6927	0.3628
Aggregated portfolio	1413	0.5194	0.3915

Table 8. Descriptions of portfolio loss distributions

Table 8 shows the characteristics of loss distributions of both aggregated and segmented portfolios. Three measurements including Value-at-Risk (VaR), expected shortfall (ES), and expected loss (EL) are reported. VaR and ES are reported at both 0.05 and 0.01 levels.

Both AIRB and FIRB approaches are examined. Under the FIRB approach the LGD of senior unsecured bond is assigned as 0.45 and the subordinated bond is assigned a value of 0.75. For the senior secured bond we use the historical average LGD of the senior secured bonds in our sample which is 0.3708 based on Table 7. Three models including obligor-varying, time-varying factor and linear regression model are considered to be AIRB approaches.

Panel A to Panel D present the statistics of aggregated and segmented portfolios, and Panel E shows t-test statistics of EL for three comparisons: 1) $\mathbf{EL}_{\textbf{Time-varying}} - \mathbf{EL}_{\textbf{Obligor-varying}}$ 2) $\mathbf{EL}_{\textbf{Linear regression}} - \mathbf{EL}_{\textbf{Obligor-varying}}$ and 3) $\mathbf{EL}_{\textbf{FIRB}} - \mathbf{EL}_{\textbf{Obligor-varying}}$, and the paired t-test statistics of ES(0.05) and ES(0.01) are reported for two comparisons including $\mathbf{ES}_{\textbf{Time-varying}} - \mathbf{ES}_{\textbf{Obligor-varying}}$ and $\mathbf{ES}_{\textbf{Linear regression}} - \mathbf{ES}_{\textbf{Obligor-varying}}$.

Panel A. Aggregated portfolio

	VaR		**ES**		**EL**
q	0.05	0.01	0.05	0.01	−
FIRB					0.5194
Obligor-varying factor model	0.6171	0.6405	0.6313	0.6515	0.5518
Time-varying factor model	0.5583	0.5823	0.5729	0.5940	0.5022
Linear regression	0.5333	0.5390	0.5368	0.5419	0.5192

Panel B. Senior secured bonds

	VaR		**ES**		**EL**
q	0.05	0.01	0.05	0.01	−
FIRB					0.3707
Obligor-varying factor model	0.4317	0.4883	0.4660	0.5158	0.2960
Time-varying factor model	0.4615	0.5128	0.4931	0.5375	0.3387
Linear regression	0.3995	0.4116	0.4068	0.4172	0.3708

Panel C. Senior unsecured bonds

	VaR		**ES**		**EL**
q	0.05	0.01	0.05	0.01	−
FIRB					0.4500
Obligor-varying factor model	0.6743	0.7080	0.6950	0.7249	0.5942
Time-varying factor model	0.5458	0.5762	0.5645	0.5918	0.4712
Linear regression	0.5099	0.5183	0.5150	0.5225	0.4900

Panel D. Subordinated bonds

	VaR		ES		EL
q	0.05	0.01	0.05	0.01	–
FIRB					0.7500
Obligor-varying factor model	0.7317	0.7487	0.7421	0.7573	0.6903
Time-varying factor model	0.7370	0.7619	0.7525	0.7752	0.6764
Linear regression	0.7189	0.7300	0.7256	0.7353	0.6927

Panel E. Paired t-test of EL, ES(0.05) and ES(0.01) at aggregated portfolio

Expected Loss				
	Mean	**Std**	**t value**	**p value**
$\mathbf{EL_{Time\text{-}varying} – EL_{Obligor\text{-}varying}}$	-0.0496	0.0002	-326.6659	<0.0001
$\mathbf{EL_{Linear\ regression} – EL_{Obligor\text{-}varying}}$	-0.0326	0.0001	-296.7424	<0.0001
$\mathbf{EL_{FIRB} – EL_{Obligor\text{-}varying}}$	-0.0324	0.0001	-304.9339	<0.0001
Expected Shortfall q =0.05				
	Mean	**Std**	**t value**	**p value**
$\mathbf{ES(0.05)_{Time\text{-}varying} – ES\ (0.05)_{Obligor\text{-}varying}}$	-0.0584	0.0002	-238.1097	<0.0001
$\mathbf{ES(0.05)_{Linear\ regression} – ES(0.05)_{Obligor\text{-}varying}}$	-0.0945	0.0002	-532.7708	<0.0001
Expected Shortfall q =0.01				
	Mean	**Std**	**t value**	**p value**
$\mathbf{ES(0.05)_{Time\text{-}varying} – ES\ (0.05)_{Obligor\text{-}varying}}$	-0.0575	0.0005	-127.5584	<0.0001
$\mathbf{ES(0.05)_{Linear\ regression} – ES(0.05)_{Obligor\text{-}varying}}$	-0.1096	0.0003	-335.9477	<0.0001